Exploring the space

- Algorithms like Metropolis-Hastings exhibit "random walk behavior" if the step size (proposal variance) is small
- If the step size is too big, then you get rejected too often
- Adaptive methods exist (see slice sampling in Bishop)
- Another approach is to combine samplers with different properties
- Another approach is to modify the distribution (e.g. annealing)

Continuous versus discrete variables

- Derivatives of continuous distributions can tell you about the structure of your problem.
  - Opportunities for going much faster
- Consider gradient ascent with added stochastic properties
  - Take a step, then perturb the result.

Hybrid Monte Carlo

- A more effective example method is Hybrid Monte Carlo
- Link the probability distribution to a potential energy function
  - Alternate stochastic sampling with "dynamics".
  - The dynamics follow the system to find low energy (high probability)
- HMC is an "auxiliary variable sampler"
  - Important trick
  - To sample p(z) we sample p(z,r) or $p(z,r_1,r_2,...)$
  - Ignore the auxiliary variables when we use the samples.

Hamiltonian Dynamics

$$p(\mathbf{z}) = \frac{1}{Z_p}\exp\left(-E(\mathbf{z})\right)$$

We equate $\mathbf{z}$ with position, so $E(\mathbf{z})$ is the potential energy.

High probability $\Leftrightarrow$ Low energy

$$E(\mathbf{z}) = -\log\left(Z_p\right) - \log\left(p(\mathbf{z})\right).$$

Hamiltonian Dynamics

Recall that the gradient, ∇, is the vector of partial derivatives.

Recall from physics that force is the negative gradient of energy

From before $E(\mathbf{z}) = -\log(Z_p) - \log(p(\mathbf{z}))$

So $\nabla E(\mathbf{z}) = \nabla(-\log(p(\mathbf{z})))$

Or, in terms of log probabilities, we define

$\Delta(\mathbf{z}) = \nabla(\log(p(\mathbf{z}))) = -\nabla E(\mathbf{z})$ (This is the force)

Hamiltonian Dynamics

$p(\mathbf{z}) = \frac{1}{Z_p}\exp(-E(\mathbf{z}))$ and $\nabla E(\mathbf{z}) = \nabla(-\log(p(\mathbf{z})))$

Let $\mathbf{r}$ be the momentum vector for the system. Denote the kinetic energy by $K(\mathbf{r})$.

$K(\mathbf{r}) = \frac{1}{2}\|\mathbf{r}\|^2 = \frac{1}{2}\sum_i r_i^2$ (We assume that mass is one).

Hamiltonian Dynamics

$H(\mathbf{z},\mathbf{r}) = E(\mathbf{z}) + K(\mathbf{r})$ (conserved)

Our distribution with auxilary variables is

$p(\mathbf{z},\mathbf{r}) = \frac{1}{Z}\exp(-H(\mathbf{z},\mathbf{r}))$

Hamiltonian Dynamics

$H(\mathbf{z},\mathbf{r}) = E(\mathbf{z}) + K(\mathbf{r})$ (conserved)

We follow $\mathbf{z}$ according to H with a random $\mathbf{r}$

This can rapidly transport us towards (but not to) a local minimun thus avoiding random walk.

To follow H, we observe that $\mathbf{z}$ changes proportional to $\mathbf{r}$, and $\mathbf{r}$ changes proportion to force ($-\nabla E$).

Again $-\nabla E = \nabla(\log(p(\mathbf{z}))) \equiv \Delta p(\mathbf{z})$

Following Dynamics

In HMC we follow the dynamics for L time steps of size τ (tunable parameters).

In the "leap frog"method for each τ.

1. Take 1/2 step in **r**.
2. Take a full step in **z**.
3. Take 1/2 step in **r**.

Following Dynamics

For L leap frog steps we have.

1. Take 1/2 step in **r**.
2. $(L-1)$ times take a full steps in **z**, then **r**.
3. Take a full step in **z**.
4. Take 1/2 step in **r**

Following Dynamics

To take a full step in **z**.

$$\mathbf{z}(\tau+1) = \mathbf{z}(\tau) + \varepsilon \cdot \Delta(\mathbf{r}(\tau))$$

(ε is the step size).

Following Dynamics

To take 1/2 step in **r**.

$$\mathbf{r}\left(\tau + \frac{1}{2}\right) = \mathbf{r}(\tau) + \frac{1}{2}\varepsilon \cdot \Delta(\mathbf{z}(\tau))$$

Where $\Delta(\mathbf{z}(\tau)) = \nabla \log\left(p(\mathbf{z}(\tau))\right) = -\nabla E(\mathbf{z})$ (force)

Following Dynamics

- After L steps of size t, we are at a new point with some bias of being at a lower potential energy (higher probability) and higher momentum.
- Momentum allows us to jump out of wells.

HMC dynamics step acceptance

- If our integration is perfect (i.e., in the limit as t --> 0) then energy is conserved.
  - Thus the value of distribution p(**z**,**r**) is the same after the dynamics.
- If we assume no integration errors, we simply accept this step
- If we want to account for error accumulation, we accept the result according to:

$$\min\left(1, \frac{p(\mathbf{z}^*, \mathbf{r}^*)}{p(\mathbf{z}, \mathbf{r})}\right) = \min\left(1, \exp\left(H(\mathbf{z}, \mathbf{r}) - H(\mathbf{z}^*, \mathbf{r}^*)\right)\right)$$

HMC stochastic step

- Typical instantiations sample the momentum variable
- Two common strategies
  - Sample the **r** independently from a Gaussian
  - Sample **r** from a Gaussian using Gibbs
- Note that in both of these cases the proposals are always accepted.

Putting it all together (A typical vision lab sampler)

- Discrete variables are sampled using (reversible jump) Metropolis Hastings.
- Continuous variables are sampled using stochastic dynamics (essentially hybrid Monte Carlo).
- Discrete variables typically control topology or components
  - The number of components and their type (block, cylinder)
  - How components are connected (branches from a stem)

A typical vision lab sampler

- Randomly proposing structure is too expensive because of the high rejection rate.
- Solution (part one) is to use data driven sampling
  - Proposals are conditioned on distributions computed before we begin using the data
  - For example, the probability of a corner being present in each point in the image.
- Solution (part two) is to delay acceptance
  - Adjust continuous parameters using stochastic dynamics so that the proposed structure is a good fit to the data.

A typical vision lab sampler

- We thus alternate between
  - (1) data driven proposals for new structure (or to switch or kill existing structure)
  - (2) exploring the continuous parameters of the structure
- Additional gains in optimization through having multiple samplers running in parallel exchange information